

# Taller GBIF.ES: Estandarización, documentación y publicación de datos de seguimientos de biodiversidad

28 – 30 de octubre de 2025, Universidad de Sevilla

## Conceptos básicos de la calidad y limpieza de datos de biodiversidad\*

Francisco Pando  
GBIF España

(\* sobre materiales de Paula Zermoglio y Katia Cezón)



# Estado de situación



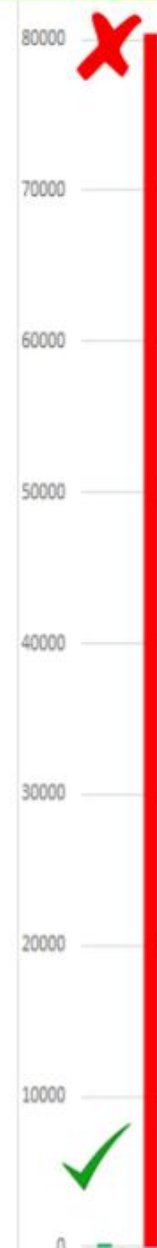
## Ejemplos:

reproductiveCondition = **40.838** valores distintos



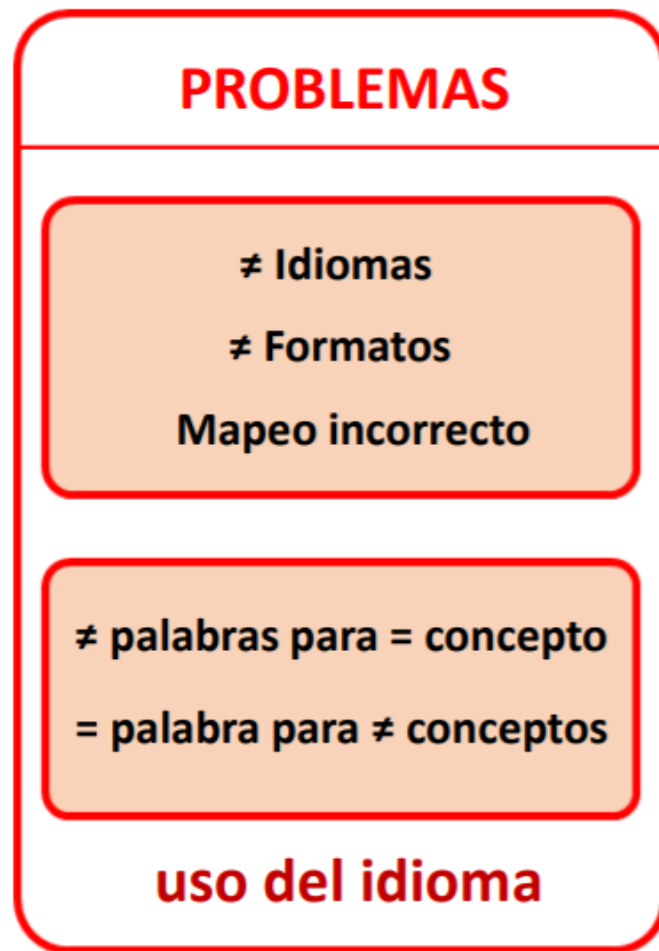
lifeStage = **33.402** valores distintos

country = **80.408** valores distintos  
(esperado = 250)





## Ejemplos:



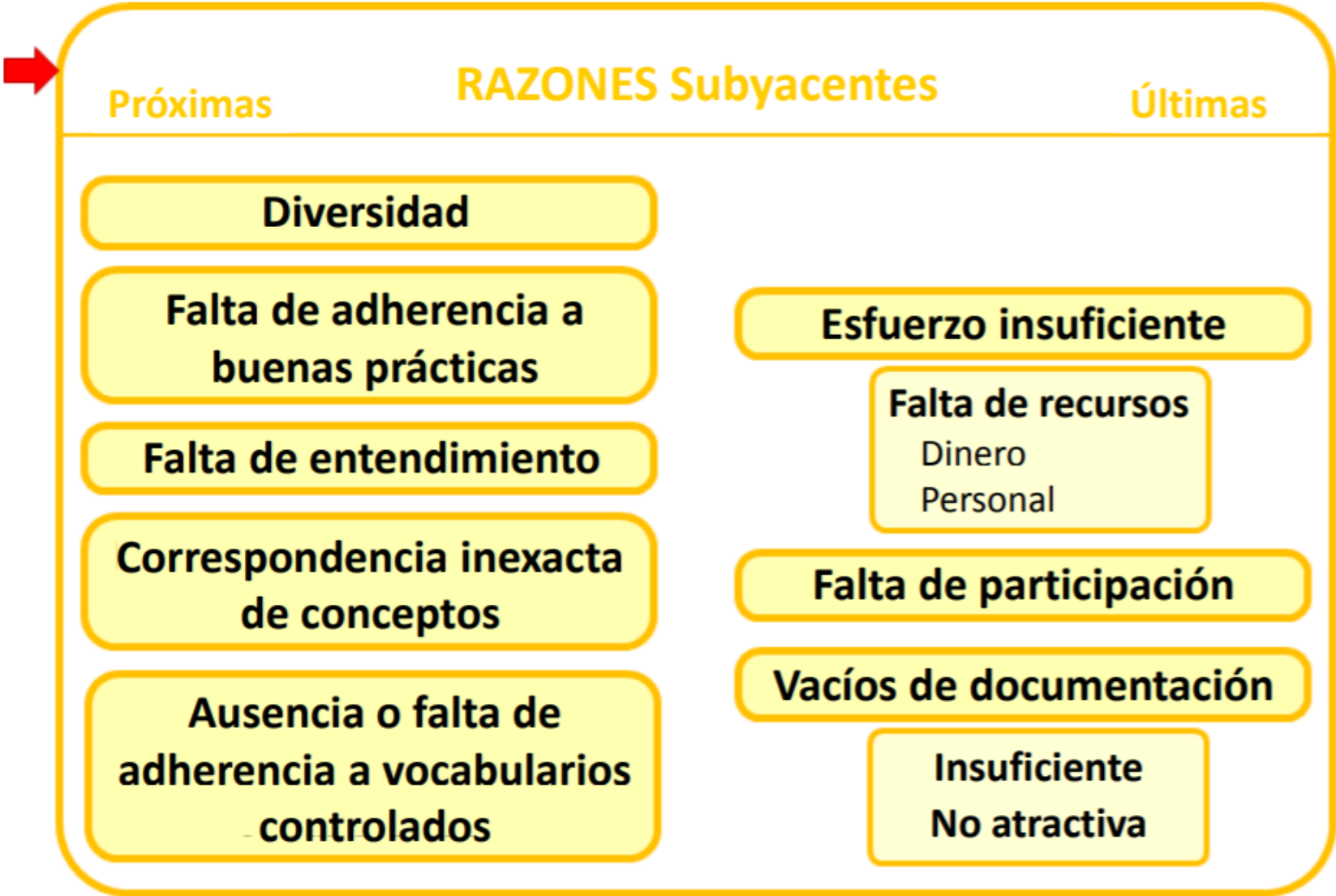
### Consecuencias

- Contenido excluído o ubicado incorrectamente
- Gran variabilidad

**Datos menos pasibles de descubrimiento y menos utilizables**



## Ejemplos:





# Impacto de la calidad de los datos

Long = 0

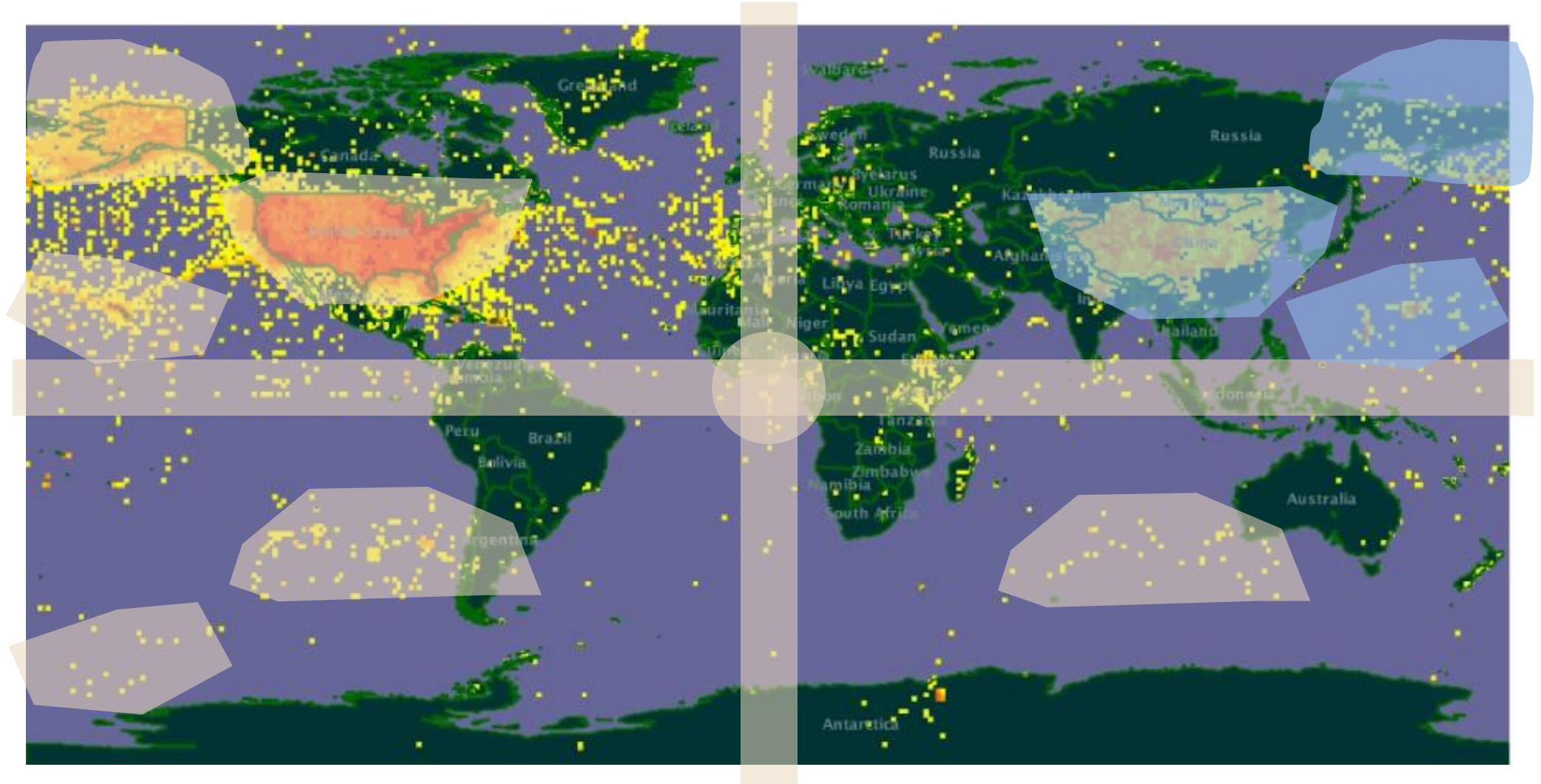
Lat = 0

Long y Lat = 0

- Long

- Lat

- Long y Lat



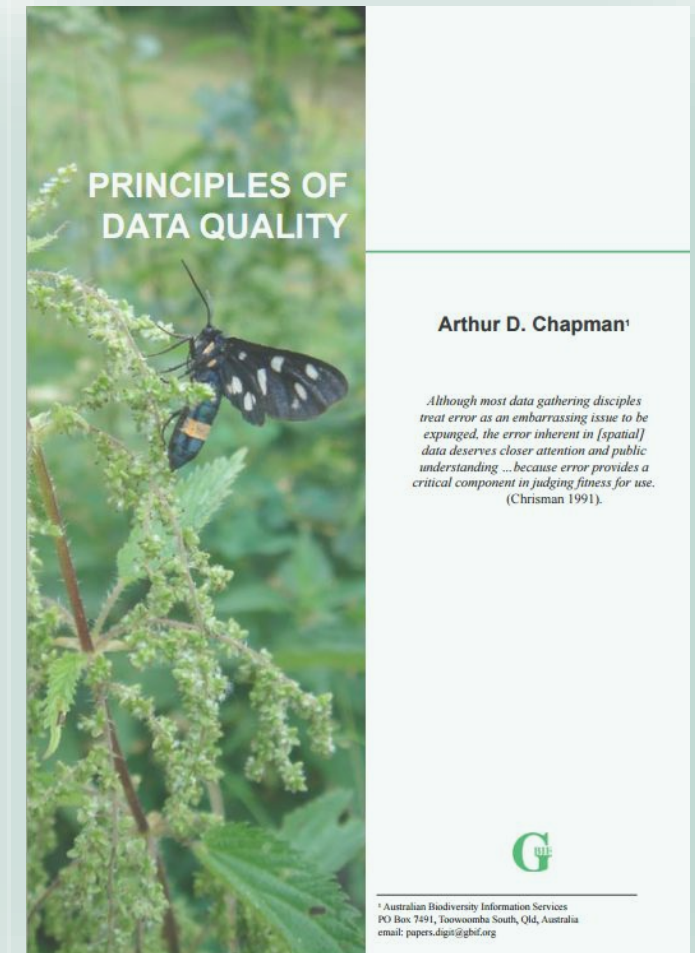
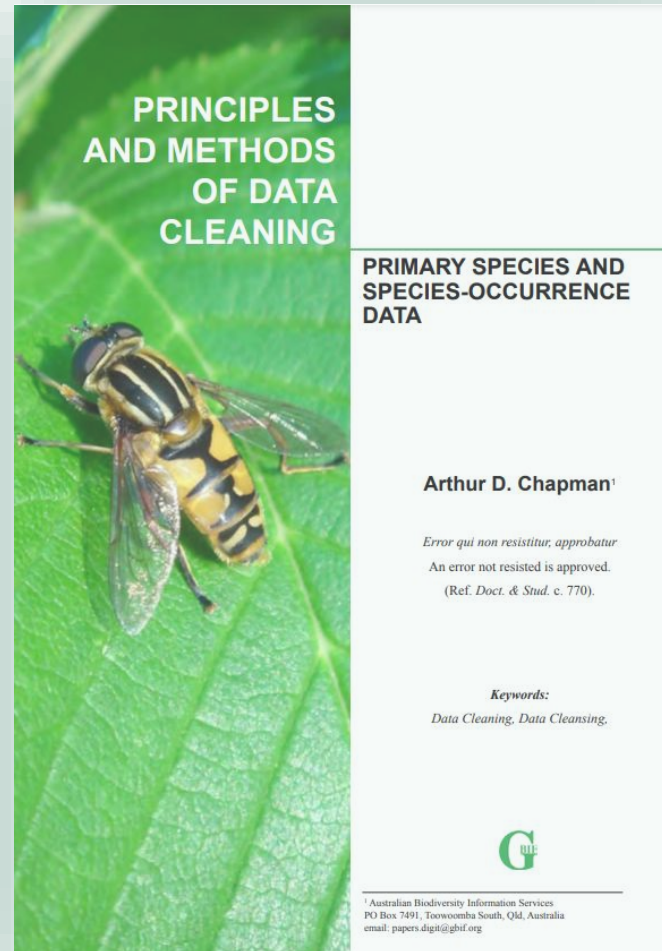
*Datos de presencia Estados Unidos GBIF (2013)*

Tim Hirsch



## Obras de referencia

- Chapman, A. D. 2005. **Principles of Data Quality**. Report for the GBIF, Copenhagen. Disponible [aquí](#).
- Chapman, A. D. 2005. **Principles and Methods of Data Cleaning – Primary Species and SpeciesOccurrence Data**. Report for GBIF, Copenhagen. Disponible [aquí](#).



# Calidad de Datos

¿Qué es la calidad de datos?

*Fitness for use = adecuación o aptitud para el uso*

La calidad de los datos está relacionada con el **uso** y no puede evaluarse independientemente del **usuario**.

La calidad está dada por el **propósito**.

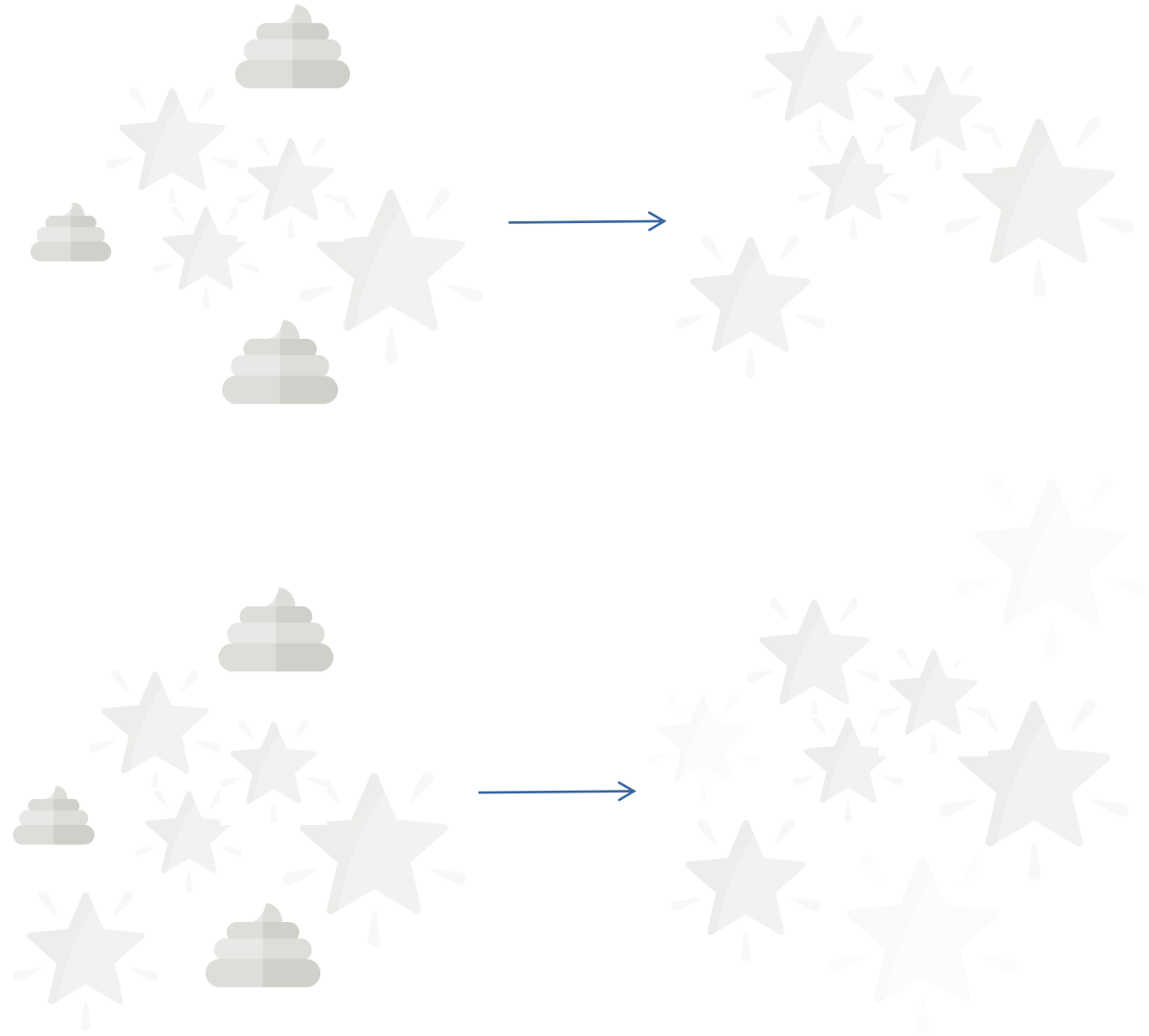
Un dato tiene calidad si es adecuado para ser usado. Si el dato no sirve para un propósito del que lo usa, puede ser útil para otros.



# Calidad de los datos

Por tanto:

Es el **usuario** es  
quien determina si  
un dato es de  
calidad o no







## ¿Qué es la calidad de datos?

**Aptitud depende de que los datos sean:**

- accesibles
- oportunos
- fáciles de leer
- fáciles de interpretar
- relevantes
- consistentes
- completos
- exactos
- detallados
- exhaustivos

# Calidad de Datos

**¿Qué es la calidad de datos?**

**Dimensiones**

**Problemas**

**Soluciones**

**Principios**

# Calidad de Datos

## Dimensiones

- ❖ **Compleitud**
- ❖ **Consistencia**
- ❖ **Exactitud**
- ❖ **Precisión**
- ❖ **Credibilidad**

# Calidad de Datos

## Problemas

- ❖ **Redundancia:** sin estándares, sinónimos.
- ❖ **Valor faltante:** valor ausente.
- ❖ **Valor incorrecto:** valor no representa los hechos.
- ❖ **Valor no atomizado:** más de un valor en un campo atomizado.
- ❖ **Esquizofrenia:** campo utilizado con un fin incorrecto.
- ❖ **Registros duplicados:** más de un registro representando el mismo hecho.
- ❖ **Valores inconsistentes:** incongruencia en la representación de un hecho.
- ❖ **Contaminación:** creación de nuevos datos a partir de datos existentes incorrectos.



# Calidad de Datos

## Soluciones

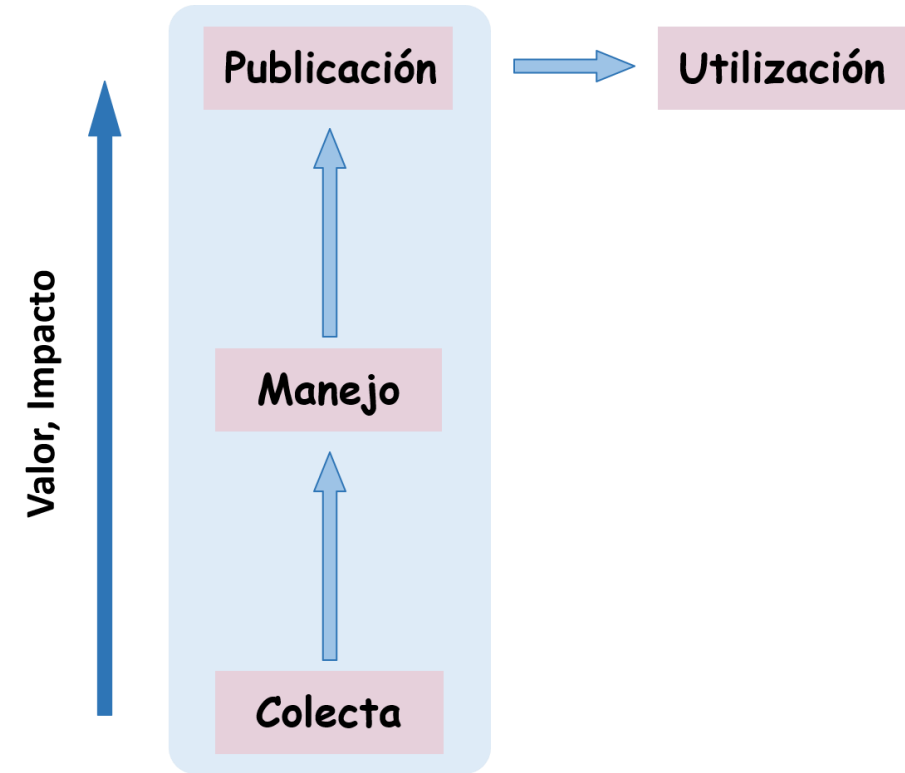
- ❖ **Prevención**
- ❖ **Detección y corrección**
- ❖ **Detección y recomendación**

# Calidad de Datos

## Soluciones

- ❖ **Prevención**
- ❖ **Detección y corrección**
- ❖ **Detección y recomendación**

Pueden actuar a diferentes niveles



# ¿Dónde surgen los problemas?

Cadena de la información



*“Ya que el error es algo ineludible, debería ser reconocido como una dimensión fundamental del dato. Y necesita ser detectado, registrado y documentado.”* [Chrisman, 1991](#).

# Mecanismos de mejora

1

## **Prevención**

Prevenir es mejor que curar (y es mucho más barato).

2

## **Detección y limpieza**

Identificar errores en el conjunto de datos y corregirlos.

3

## **Documentación y recomendaciones**

Detectar errores en el conjunto de datos y generar recomendaciones de limpieza.

4

## **Mecanismos de retroalimentación**

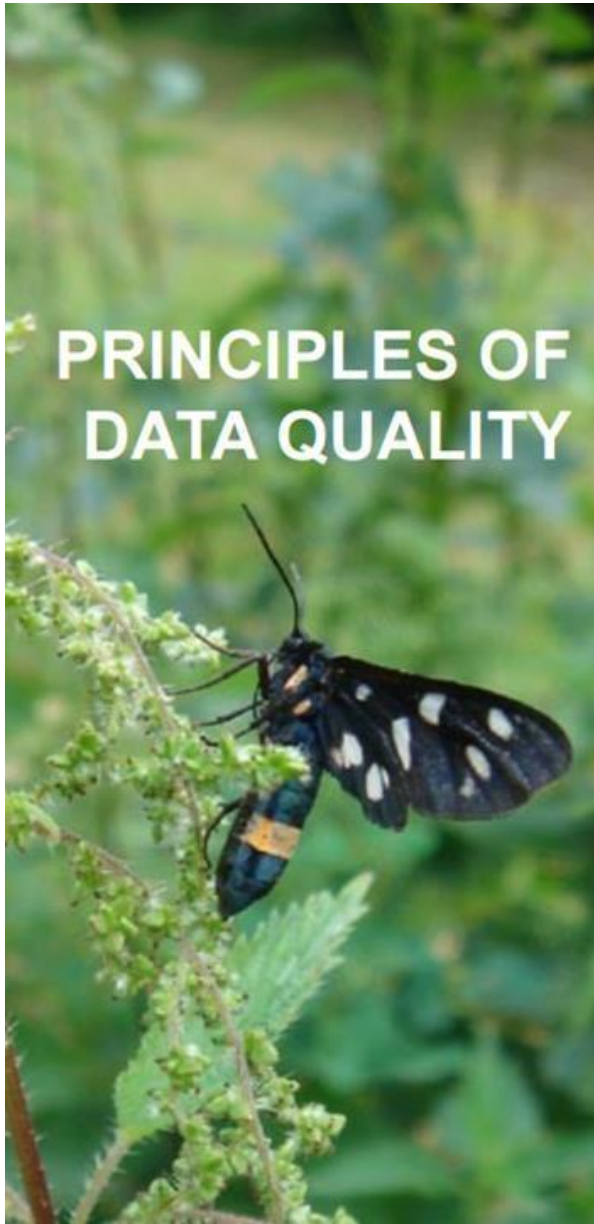
Informar de los errores a los actores implicados en los pasos previos.

5

## **Capacitación**

Del personal que colecta y manipula datos





## Principios

- ❖ **Planeamiento**
- ❖ **Organización**
- ❖ **Prevención**
- ❖ **Responsabilidad**
- ❖ **Asociación**
- ❖ **Priorización**
- ❖ **Medidas de desempeño**
- ❖ **Optimización**
- ❖ **Retroalimentación**
- ❖ **Capacitación**
- ❖ **Transparencia**
- ❖ **Documentación**

# Calidad de Datos

## Principios

- No podemos ir atrás en el tiempo... el potencial nunca será tan alto como en el punto de origen.

# Calidad de Datos

## Principios

**Capturar tanto y tan inequívocamente  
como sea posible.**

# Calidad de Datos

## Principios

**Capturar tanto y tan inequívocamente  
como sea posible.**

- La falta de conocimiento puede resultar en potenciales errores.



# Calidad de Datos

## Principios

**Capturar tanto y tan inequívocamente  
como sea posible.**

**Capacitar al personal que colecta  
y manipula datos.**

# Calidad de Datos

## Principios

**Capturar tanto y tan inequívocamente  
como sea posible.**

**Capacitar al personal que colecta  
y manipula datos.**

- Toda transformación de los datos puede potencialmente resultar en una pérdida neta de información.

# Calidad de Datos

## Principios

**Capturar tanto y tan inequívocamente como sea posible.**

**Capacitar al personal que colecta y manipula datos.**

**Mantener todos los datos originales.**

**Documentar todos los cambios hechos.**

# Calidad de Datos

## Principios

**Capturar tanto y tan inequívocamente  
como sea posible.**

**Capacitar al personal que colecta  
y manipula datos.**

**Mantener todos los datos originales.**

**Documentar todos los cambios hechos.**

- El uso de los datos aporta al control de calidad.



# Calidad de Datos

## Principios

**Capturar tanto y tan inequívocamente como sea posible.**

**Capacitar al personal que colecta y manipula datos.**

**Mantener todos los datos originales.**

**Documentar todos los cambios hechos.**

**Asegurar que los datos y la documentación sean accesibles.**

**Utilizar mecanismos de retroalimentación desde los usuarios.**

# Calidad de Datos

## Ejemplo final: ¿cómo usar este registro?



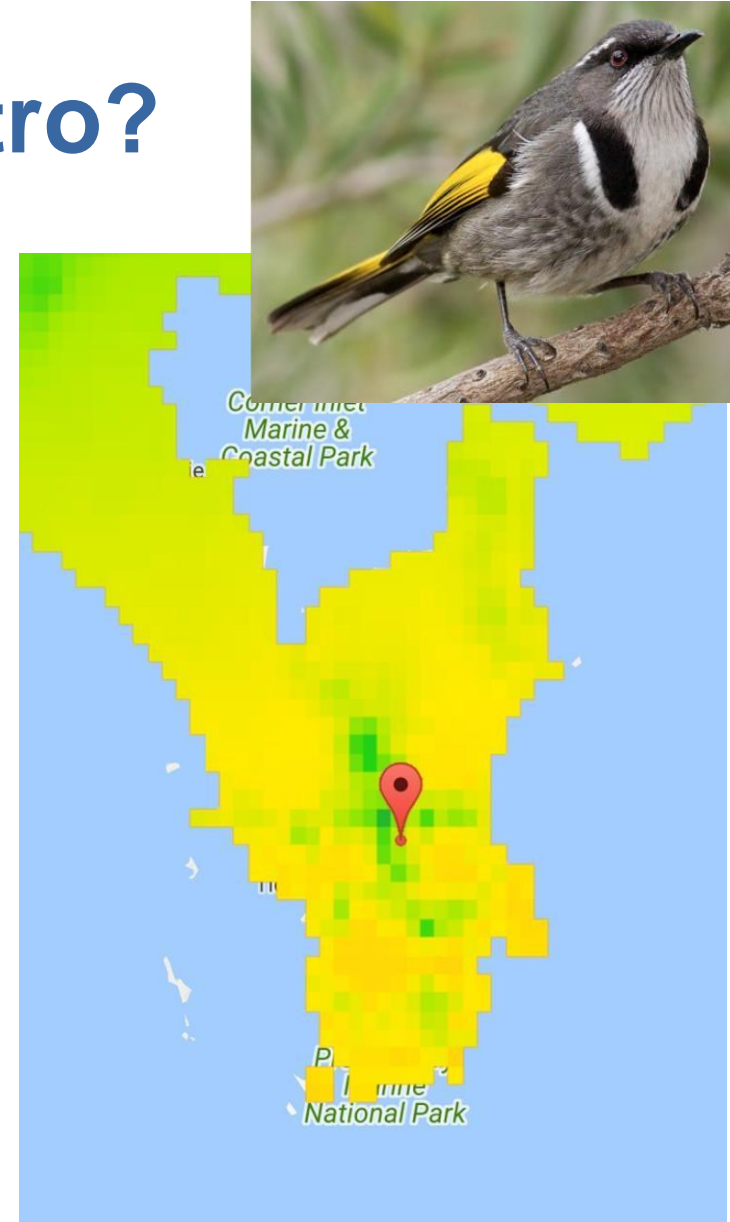
-38.99373,  
146.37367

Elevación: 302m

Temp., mes más frío:

6.1°C Precip., media

anual: 107mm



# Calidad de Datos

## Ejemplo final: ¿cómo usar este registro?



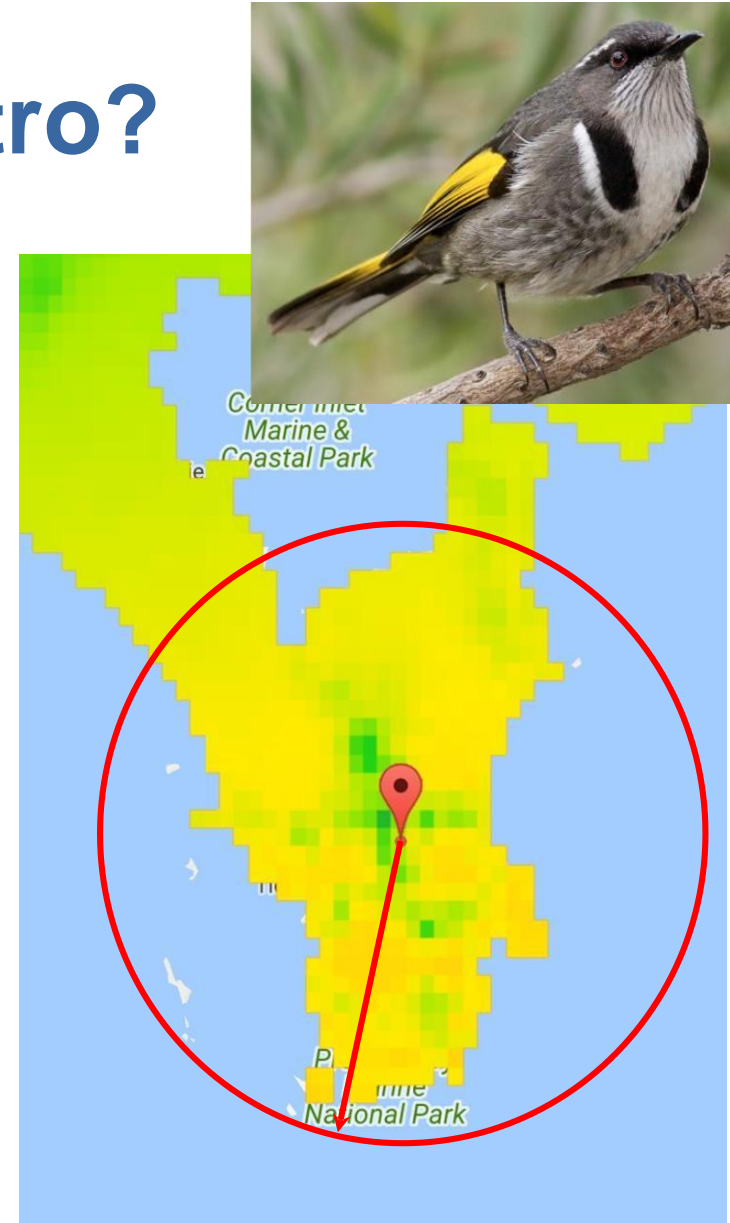
-38.99373, 146.37367  
(Wilson's Promontory)  
coordinateUncertainty: 14570m

Elevación: 0-544m (302m)

Temp., mes más frío: 6.1-9.1°C (6.1°C)

Precip., media anual: 65-107mm

(107mm)



# Calidad de Datos

La calidad de los datos determina su aptitud para usos particulares.

La calidad de los datos condiciona la calidad de la investigación y la toma de decisiones.

Francisco Pando

Unidad de Coordinación de GBIF.ES

CSIC

Joaquín Costa 22, 28002 Madrid

[pando@gbif.es](mailto:pando@gbif.es)



<https://creativecommons.org/licenses/by/4.0/deed.es>

GBIF-ES es el Nodo Nacional de Información en Biodiversidad patrocinado por el Ministerio de Ciencia, Innovación y Universidades, gestionado por el Consejo Superior de Investigaciones Científicas (CSIC). <https://www.gbif.es/> <https://datos.gbif.es>, <https://elearning.gbif.es>



MINISTERIO  
DE CIENCIA, INNOVACIÓN  
Y UNIVERSIDADES



**CSIC**  
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

